07 - Confidence Regions for the Mean

Junvie Pailden

SIUE, F2017, Stat 589

September 10, 2017

Confidence Regions for μ

- The set of of all μ satisfying this inequality form an ellipsoid.
- For $p > 3$, this is hard to visualize and so this equality is of more mathematical interest than of practical use.
- The hypothesized mean value μ_0 lies within the confidence region if the computed the generalized square distance satisfies

$$n(\bar{\mathbf{X}} - \mu_0)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu_0) \leq \frac{(n-1)p}{(n-p)} F_{p.n-p}(\alpha)$$

- This approach is analogous to testing

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_0 : \mu \neq \mu_0$$

where T^2-test would not reject H_0 when

$$T^2 \leq \frac{(n-1)p}{(n-p)} F_{p.n-p}(\alpha).$$

Simultaneous Confidence Statements for $\mathbf{a}'\mu$

- A $100(1-\alpha)\%$ simultaneous confidence intervals involving the $(p \times 1)$ vector $\mathbf{a}$ for $\mathbf{a}'\mu$ is

$$\mathbf{a}'\bar{\mathbf{X}} - \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}} \leq \mathbf{a}'\mu \leq \mathbf{a}'\bar{\mathbf{X}} + \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

- The Simultaneous T^2 confidence intervals for all $\mathbf{a}'\mu$ are just the shadows (projection), of the confidence ellipsoid on the component axes.

In particular, if we let $\mathbf{a}' = [0, \ldots, 0, 1, 0, \ldots, 0]$ where 1 is on the ith row of $\mathbf{a}$, then a $100(1-\alpha)\%$ confidence interval for $\mathbf{a}'\mu = \mu_i$ $(p = 1)$ is

$$\bar{X}_i - \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{X}_i + \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{ii}}{n}}$$

$$\bar{X}_i - \sqrt{F_{1,n-1}(\alpha)} \sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{X}_i + \sqrt{F_{1,n-1}(\alpha)} \sqrt{\frac{s_{ii}}{n}}$$

$$\bar{X}_i - t_{n-1}(\alpha/2) \sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \bar{X}_i + t_{n-1}(\alpha/2) \sqrt{\frac{s_{ii}}{n}}$$

Simultaneous Confidence Statements for $\mathbf{a}'\mu$

- We can also make statements about the differences $\mu_i - \mu_k$ corresponding to $\mathbf{a}' = [0, \ldots, 0, a_i, 0, \ldots, a_k, \ldots, 0]$, where $a_i = 1$ and $a_k = -1$. In this case $(p = 2)$, $\mathbf{a}'\mathbf{S}\mathbf{a} = s_{ii} - 2s_{ik} + s_{kk}$, we have the interval

$$(\bar{X}_i - \bar{X}_k) - \sqrt{\frac{(n-1)p}{(n-p)} F_{p.n-p}(\alpha)} \sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}} \leq \mu_i \leq$$

$$(\bar{X}_i - \bar{X}_k) + \sqrt{\frac{(n-1)p}{(n-p)} F_{p.n-p}(\alpha)} \sqrt{\frac{s_{ii} - 2s_{ik} + s_{kk}}{n}}$$

- Which set of intervals is better (smaller) depends on the relative sizes of n and p, and even the number of means compared, say m μ_i's.
- Perhaps the best way is to calculate both sets (critical values) and use the set yielding the narrower intervals.

Patients Example : Simultaneous Statenents

Test to see if $\mu_2 = \mu_5$.

$$H_0 : \mu_2 = \mu_5 \text{ vs } H_1 : \mu_2 \neq \mu_5$$

```
patients <- read.csv("patients.csv", header=TRUE)
(Xbar <- colMeans(patients))
```

```
#   WEIGHT FASTING GLUCOSE INSULIN  RESIST
#     0.92   90.41  340.83  171.37   97.78
```

```
S <- cov(patients) # cov matrix
n <- nrow(patients)
```

We want to see if $\mu_2 = \mu_5$.

$$H_0 : \mu_2 = \mu_5 \text{ vs } H_1 : \mu_2 \neq \mu_5$$

```r
p <- 2 # comparing two means
(cval <- ((n-1)*p/(n-p))*qf(0.95,
                       df1 = p, df2 = n-p)) # F critical value
```

```
# [1] 6.6
```

```r
a <- c(0,1,0,0,-1)
Ybar <- t(a)%*%Xbar
SY <- t(a)%*%S%*%a
LL <- Ybar - sqrt(cval)*sqrt(SY/n)
UL <- Ybar + sqrt(cval)*sqrt(SY/n)
data.frame(Mean.D = Ybar, Lower.lim = LL, Upper.lim = UL)
```

```
#   Mean.D Lower.lim Upper.lim
# 1   -7.4       -25        11
```

The 95% simultaneous confidence interval for mu2-mu5 is (-25, 11). Since 0 is inside the confidence interval, then it is plausible that H0 holds.

Bonferroni Method of Multiple Comparisons

- Suppose prior to the collection of data, confidence statementts about m linear combinations $\mathbf{a}_1'\mu, \mathbf{a}_2'\mu, \ldots, \mathbf{a}_m'\mu$ are required.
- Let C_i denote the confidence statement abou the value $\mathbf{a}_i'\mu$ with $P(C_i \text{ true}) = 1 - \alpha_i, i = 1, 2, \ldots, m$.

$$P[\text{all}\, C_i \,\text{true}] \geq 1 - \sum_{i=1}^{m} \alpha_i$$

- A special case of the Bonferroni allows the investigator to control the overall error rate $\sum_{i=1}^{m} \alpha_i$, regardless of the correlation structure.
- We consider the individual t-intervals

$$\bar{X}_i \pm t_{n-1}\left(\frac{\alpha_i}{2}\right)\sqrt{\frac{s_{ii}}{n}}, \;\; i = 1, 2, \ldots, m,$$

where $\alpha_i = \alpha/m$.

```
p <- 5; alpha <- 0.05
tci <- qt(1-alpha/2,df=n-1)
bc.tci <- qt(1-alpha/(2*p),df=n-1)
T2ci <- sqrt(((n-1)*p/(n-p))*qf(0.95, df1=p, df2=n-p))
CI <- function(cval,Xbar,S,n){
cbind(Xbar-cval*sqrt(diag(S/n)),
      Xbar+cval*sqrt(diag(S)/n))}
# t-intervals
t <- CI(tci, Xbar, S, n)
# bonferroni corrected t-intervals
bct <- CI(bc.tci,Xbar,S,n)
# Hotelling T2-intervals
T2 <- CI(T2ci,Xbar,S,n)
```

```r
# Confidence Intervals for the mean vector with
# alpha = 0.05
# One-at-a-time t, bonferroni corrected t,
# and T2 intervals
colnames(t) <- colnames(bct) <- colnames(T2) <- c("LL", "UI
data.frame(t = t, Bonf_t = bct, T2 = T2)
```

```
#            t.LL   t.UL Bonf_t.LL Bonf_t.UL  T2.LL  T2.UL
# WEIGHT     0.88   0.96      0.87      0.97   0.85   0.99
# FASTING   87.92  92.91     87.08     93.74  85.88  94.95
# GLUCOSE 330.95 350.70    327.64    354.02 322.87 358.78
# INSULIN 156.88 185.86    152.02    190.72 145.02 197.72
# RESIST   84.06 111.51     79.45    116.11  72.83 122.74
```

Observations

- In general, the width of T^2-intervals, relative to t and bonferroni corrected t intervals, increases as p increases (for fixed n) and decreases as n increases (for fixed p).
- The confidence level associated with any collection of T^2-intervals, for fixed n and p, is $1 - \alpha$, and the overall confidence associated with a collection of t intervals, for the same n, can be much less.
- The bonferroni correction guarantees that overall confidence level is greater than or equal to 0.95.
- Because Bonferroni correction is easy to apply and provide relatively short confidence intervals; it often used in practice.